



science and policy  
for a healthy future

# Data in HBM4EU

Eva Govarts, Sylvie Remy, Liese Gilles,  
Greet Schoeters (VITO)

Developed in close collaboration with  
JRC (Silvia Dalla Costa)

Advanced training course on data  
management

21/06/2018, Ljubljana

# Collaboration-agreement IPCHEM – HBM4EU



The **European Human Biomonitoring Initiative (HBM4EU)** is a joint effort of 28 countries and the European Commission, co-funded by Horizon 2020. The main aim of the initiative is to coordinate and advance human biomonitoring in Europe.

## IPCHEM support:

- Share existing HBM data
- Share new HBM data
- Enhanced data and metadata model and functionalities
- Dissemination of results
- Long-term repository
- Cross-media analysis
- Input and publication of Indicators

21/06/2018



# Collaboration-agreement IPCHEM – HBM4EU

## IPCHEM-HBM4EU Collaboration

IPCHEM team defined and set up the HBM4EU Data Repository, ready to be used by HBM4EU data providers and partners

- **Secure Environment** → Security Assessment passed in May/June 2017)
- **Data can be shared only to and across authorised user accounts**
- **“User-centric” Interface** → you see only the files and folders created by and/or shared with you)

HBM4EU WP 10 on Data management and IPCHEM team defined the HBM4EU-IPCHEM metadata template, filled/to be filled by the Data Providers

IPCHEM prepared the web interfaces for the HBM4EU-IPCHEM metadata, revised by HBM4EU WP 10 and some data providers

IPCHEM team designed and developed the advanced search functionalities, based on feedback received by HBM4EU WP 10 and some data providers

Some activities to publish HBM4EU indicators are planned for the end of 2018 and in 2019, with WP 5.4 (involving EEA)

21/06/2018



# What is the purpose of the data collection and its relation to the objectives of HBM4EU?

The purpose is to **collect** and **harmonize HBM data** across Europe. The data collection will comply with all national and EU ethics and legal requirements. **Access to use these data is needed to address the different HBM4EU objectives** as specified in the grant agreement and summarized below:

Objective 1: Laying the foundations for a pan-European HBM platform that builds on national hubs and existing expertise

Objective 2: Developing a common methodology for the interpretation and use of HBM data in policy-making

Objective 3: Harmonising and optimising the practices of national HBM programmes, including sample collection, quality assurance and data management

Objective 4: Identifying gaps where further data are needed to inform current policy questions and design new, targeted studies to address these knowledge gaps

Objective 5: Including new HBM data and, where possible, existing HBM data in the European Commission's Information Platform for Chemical Monitoring (IPChem, <https://IPCHEM.jrc.ec.europa.eu/>)

# What is the purpose of the data collection and its relation to the objectives of HBM4EU?

Objective 6: Linking external to internal exposure in order to improve exposure models for risk assessment

Objective 7: Developing, validating, and applying exposure and effect biomarkers to improve our understanding of the health risks associated with aggregate exposures

Objective 8: Identifying chemicals of concern through novel methods for the holistic analysis of HBM samples and improving the use of HBM data in assessing exposure to and the risks of chemical mixtures

Objective 9: Enhancing our understanding of the causal association between chemical exposure and adverse health outcomes by combining mechanistic studies with existing cohort data

Objective 10: Promoting capacity building at national level through training and exchange programmes

Objective 11: Engaging with stakeholders, including the general public, throughout the programme to ensure the credibility, accountability and legitimacy of activities and results

# Data management in HBM4EU

To solve the Research Questions we will need:

Data collections on human subjects across Europe which contain **Personal information** and **health information** about the study participants

When **handling personal data** we have to be **compliant with the General Data Protection Regulation** (GDPR\*, Regulation (EU) 2016/679)

*\*The GDPR is a regulation by which the European Parliament, the European Council and the European Commission intend to strengthen and unify data protection for individuals within the European Union (EU).*

# General Data Protection Regulation

## Important aspects of the GDPR:

- The **pseudonymation and/or encryption** of personal data.
- The ability to **restore the availability and access to data in a timely manner in the event of a physical or technical incident.**
- A process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for **ensuring the security of the data processing.**

# FAIR data

- **F**indable
  - **Making data findable**, including providing metadata:  
-> gain visibility for your dataset to the outside world!
- **A**ccessible
  - Specify which data will be made openly available?
  - If some data is kept closed provide rationale for doing so
  - Specify how access will be provided in case there are any restrictions
- **I**nteroperable
  - Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability
- **R**eusable
  - Specify how the data will be licenced to permit the widest reuse possible
  - specify why and for what period a data embargo is needed



# Challenge, keeping the balance:

## OPEN ACCESS



**FAIR** data:  
Findable, **A**ccessible, Interoperable, **R**eusable

## GDPR



Data on human subjects  
Health data (**sensitive** data)

***“As open as possible, as closed as necessary”***

# Anonimysation vs. Pseudonimysation

Mostly HBM data originate from longitudinal studies/cohorts or cross-sectional studies:

Identification keys are typically NOT destroyed by data owner to encertain follow-up, or to communicate results to participants

→ *This means your data is pseudonimysed data*

Risk of re-identification by combining variables and datasets (especially in smaller datasets risk is higher)

→ *This means your data is pseudonimysed data*

→ Data that are shared with HBM4EU will (most likely) be pseudonymised, up to the level as to not interfere with the quality of the research.

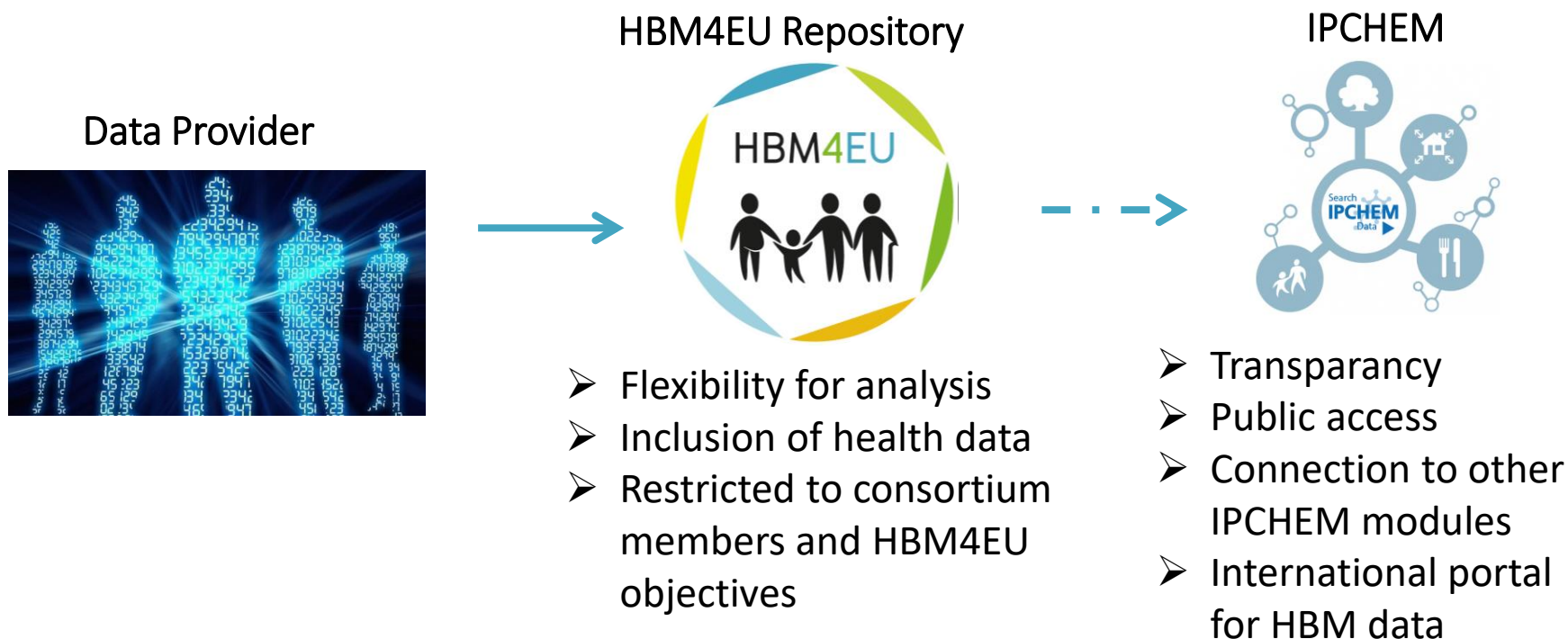
# How will we facilitate transfer of personal data in a GDPR compliant way:

- Set-up a **neutral/secure environment for hosting HBM data** (=HBM4EU repository)
  - ↳ Avoid transfer of data via e.g. e-mail and commercial cloud services (making security of the data dependent on the service provider)
- **Improve transparency** of data sharing:
  - Incorporate **to grant and revoke specific access rights**, e.g. different rights for statistical working group as compared to consortium member not involved in the analysis; compound specific access for substance specific researchers; ...
  - **Logging of data access, download, and upload**
- **Harmonization of data**: Generation of aggregated data to ensure comparability of results and hence to draw clear-cut conclusions
  - Use of templates and predefined analysis scripts has been incorporated in the data-flow.

# HBM4EU repository

= neutral/secure platform for hosting HBM data

→ HBM4EU repository will have the same security level as IPCHEM and will be fully aligned with the: IPCHEM Data Policy and HBM4EU Data Policy



# HBM4EU repository:

## neutral/secure platform for hosting HBM data

The HBM4EU repository hosted at JRC:

- A platform that enables sharing of data, results, analysis pipelines, intermediate results;
- A component of the IPCHEM architecture and allows integrating part of data from the repository into IPCHEM, !!when agreed by the Data Owners/Data Providers
- Aims to reach the highest level of GDPR compliancy by:
  - Relying on the EU authentication platform and security protocols for data sharing.
  - Applying a strict policy in granting and revoking specific access rights to the data.
  - Logging of user identity during data access, download, and upload, including version control.

# HBM4EU repository:

## neutral/secure environment for hosting HBM data

### Improve transparency:

- Allows logging of data access, download, and upload, including version control.
- Data Owners/Data Providers can chose to which research they will contribute with their data according to purpose and interests of the Data Owners/Data Providers

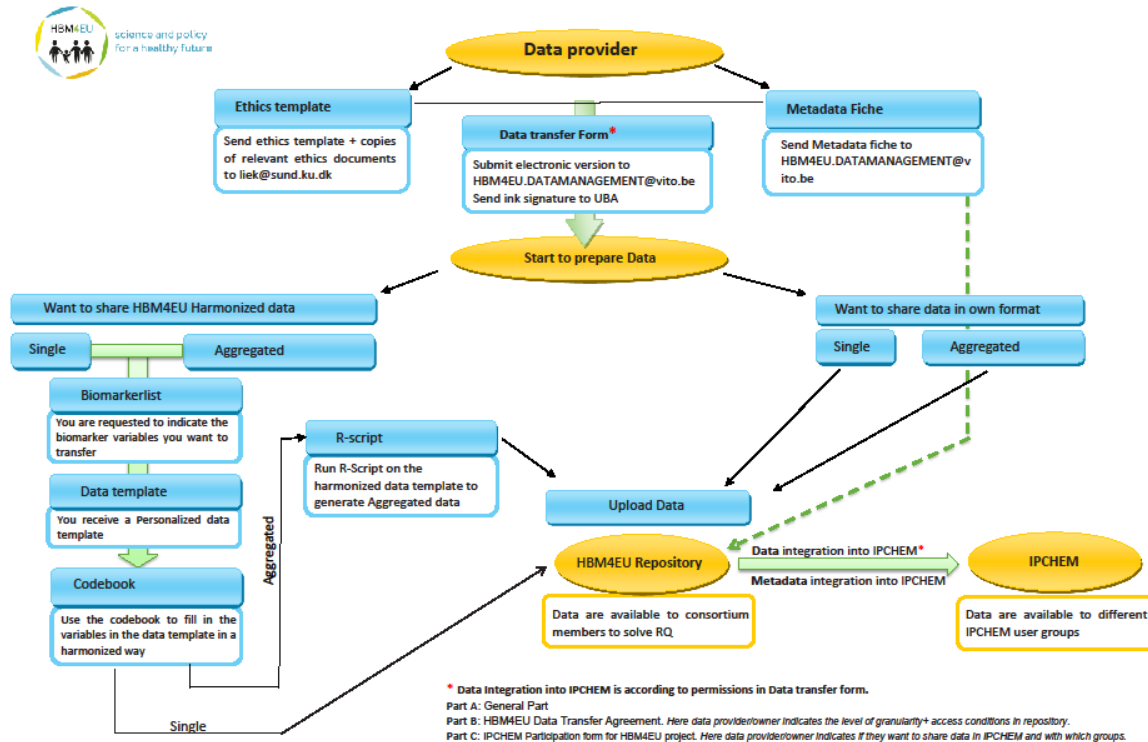
# Harmonisation of the data

- WP10 developed a **codebook** to harmonise variables
- Use an **excel template** (provided by WP10) to import the **data in a structured/harmonized manner**;
- Use an **R-script** (provided by WP10) to generate harmonized (stratified) aggregated data

*(it will be verified whether R provides all necessary analysis functions, such as weighted calculations; if not, complementary software syntax may be considered)*

-> more detailed information later

# Process of sharing data: What are the steps for data providers?





Process of sharing data:  
What are the steps for data providers?

Ethics

Metadata

Data transfer agreement (DTA)

Part B: data transfer agreement repository

Part C: IPCHEM participation form

# Ethics

- Ethics template
- Copy of informed consent
- Copy of approval
- Copy of data protection approval (if applicable)
- ...

-> explained during basic training and advanced ethics course by Lisbeth Knudsen

# Metadata

“a set of data that describes and gives information about other data”  
Collects information about your dataset NOT actual data



Microsoft Excel  
Worksheet

Metadata template



Microsoft Excel  
Worksheet

Metadata with example metadata from Flemish environment  
and health study –adolescents (FLEHS III)

# Data transfer agreement

HBM4EU Data Transfer Form	
Purpose of the form and signatures	P. 1

Version Number of the HBM4EU Data Transfer Form:

Release Date:

## 1) Purpose of this form:

The data transfer documentation is to be completed by Data Owners/Data Providers. In case a Data Provider is assigned by the Data Owner, the Data Provider will be responsible for implementation of the procedures described in the HBM4EU data management plan and data policy.

Using this form, Data Owners/Data Providers indicate under which conditions the data to which this agreement applies ([Part A](#)) are transferred to the HBM4EU repository ([Part B](#)) and integrated into IPCHEM ([Part C](#)). The metadata shall be made openly available via IPCHEM as minimal requirement.

## 2) Signatures:

Signing this form the Data Owner/Data Provider:

1) agrees to make the data collection(s) (specified in [Part A](#) of this form),

- accessible through the HBM4EU repository, under the conditions specified in [Part B](#) of this form, and accepts the HBM4EU Data Policy,
- accessible through the IPCHEM platform, under the conditions specified in [Part C](#) of this form, and accepts the IPCHEM Data Policy.

2) declares that transfer of the data to HBM4EU under the conditions specified in this form is allowed from ethico-legal perspective and agrees to collect and provide the Ethics documents. This includes, but is not limited to, documenting signed consent forms of participants and ethics approval that grant use within HBM4EU. Instructions are available at <https://c.deic.dk/hbm4euethics>. The Ethics documents shall be sent to task 1.5 leader ([liek@sund.ku.dk](mailto:liek@sund.ku.dk)) and the coordinator ([HBM4EU@uba.de](mailto:HBM4EU@uba.de))

3) agrees that the transfer of data to the HBM4EU repository and eventually the integration into IPCHEM are not allowed prior to providing this signed Data Transfer Agreements (this form), the Ethics documents and the filled out metadata template (to be sent via e-mail to [HBM4EU.DATAMANAGEMENT@vito.be](mailto:HBM4EU.DATAMANAGEMENT@vito.be)).

# Rules for submitting data to HBM4EU repository and IPCHEM

- Provide [ethical documentation to prove ethical compliance!](#)
- All personal data on human subjects that are transferred to the HBM4EU repository need to be either [pseudonymised or completely anonymized](#), as described in the HBM4EU data policy.
- Data provider [will not transfer identifiable variables](#). Directly identifiable variables include - but are not limited to:
  - national ID number
  - name
  - phone number
  - ZIP-code
  - e-mail address
  - address (including GEO-code at a resolution that risks re-identification)
- Data provider has to specify conditions under which they agree to make their data accessible for research within HBM4EU
- The HBM4EU coordinator declares that the **data that are transferred will be stored in the HBM4EU repository for the course of the project (until 31/12/2021);** and that the **data in the HBM4EU repository are managed by the procedures described in the HBM4EU data policy**

# Conditions for: Existing data –non HBM4EU funded

## Existing data - non co-funded data

*For existing data, not generated with HBM4EU co-fund, the **Data Owner/Data Provider specifies the level of granularity** that data will be transferred as*

*1. Anonymised single measurement data*

*Or*

*2. Pseudonymised single measurement data*

*Or*

*3. Aggregated data*

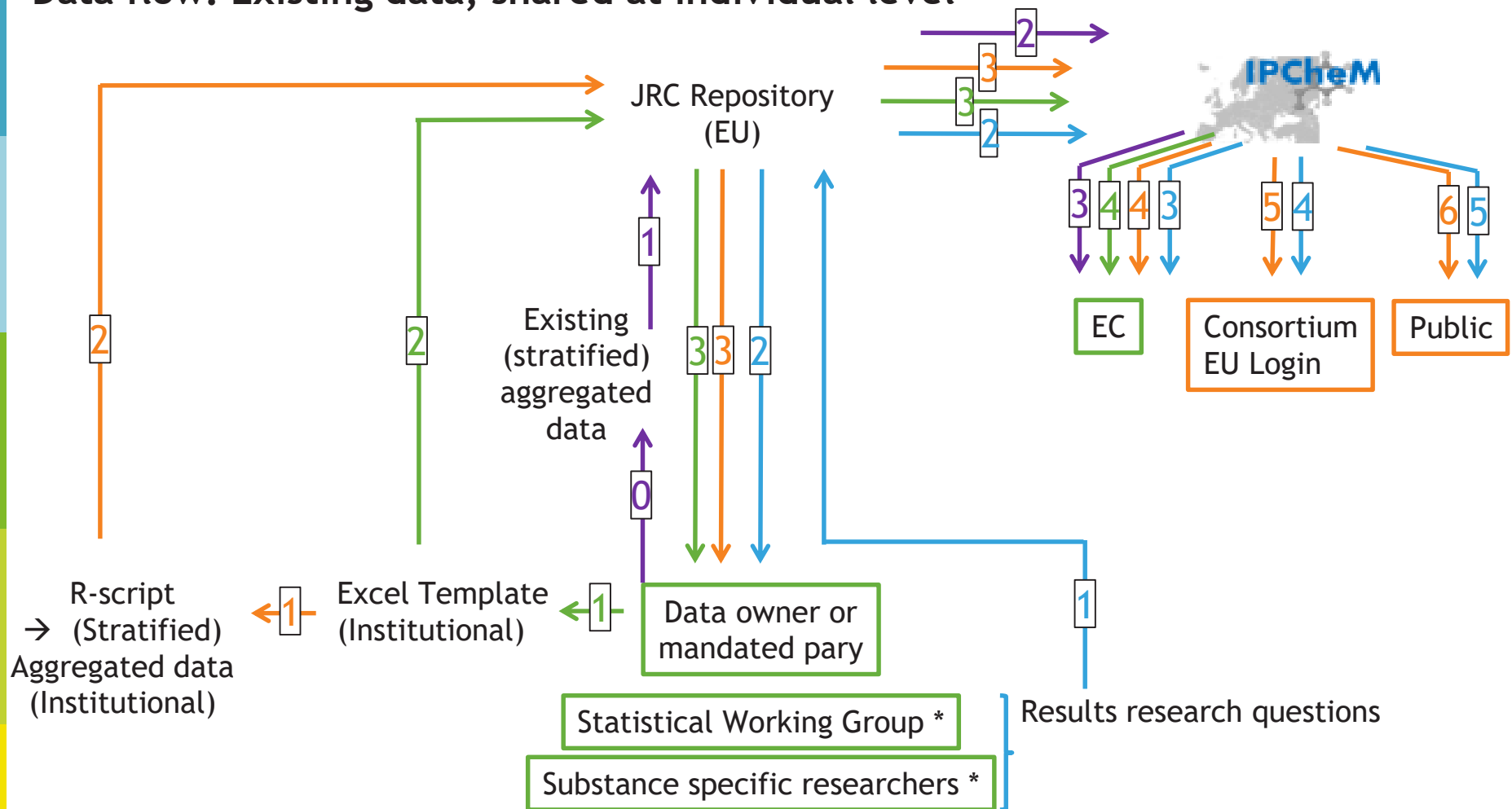
*The **Data Owner/Data Provider indicates** for each level of granularity **whether the data are:***

***directly accessible** for use within HBM4EU or*

***NOT directly accessible** for use within HBM4EU*

*Not directly accessible →the **Data Owner/Data Provider will be asked approval when consortium members request access to the data to meet the goals of a particular objective.***

# Data flow: Existing data, shared at individual level



Existing aggregated data

Resolution of the data: Individual data

Aggregated data

Results research questions HBM4EU

\* Specified users may be granted/revoked specified access rights (e.g. to variables that are needed to answer research question)

# Conditions for: Data –generated with HBM4EU co-fund

## EU Co-funded data

*For data generated with HBM4EU co-fund during the course of HBM4EU, the **Data Owner/Data Provider must agree** that these data are transferred as **single measurement data** (anonymised or pseudonymized) to the **HBM4EU repository** (and IPCHEM)*

*Accompanying variables (i.e. age, sex, BMI, smoking status etc.) that are needed to solve the envisaged **research purpose(s)** are also provided as **single measurement data**.*

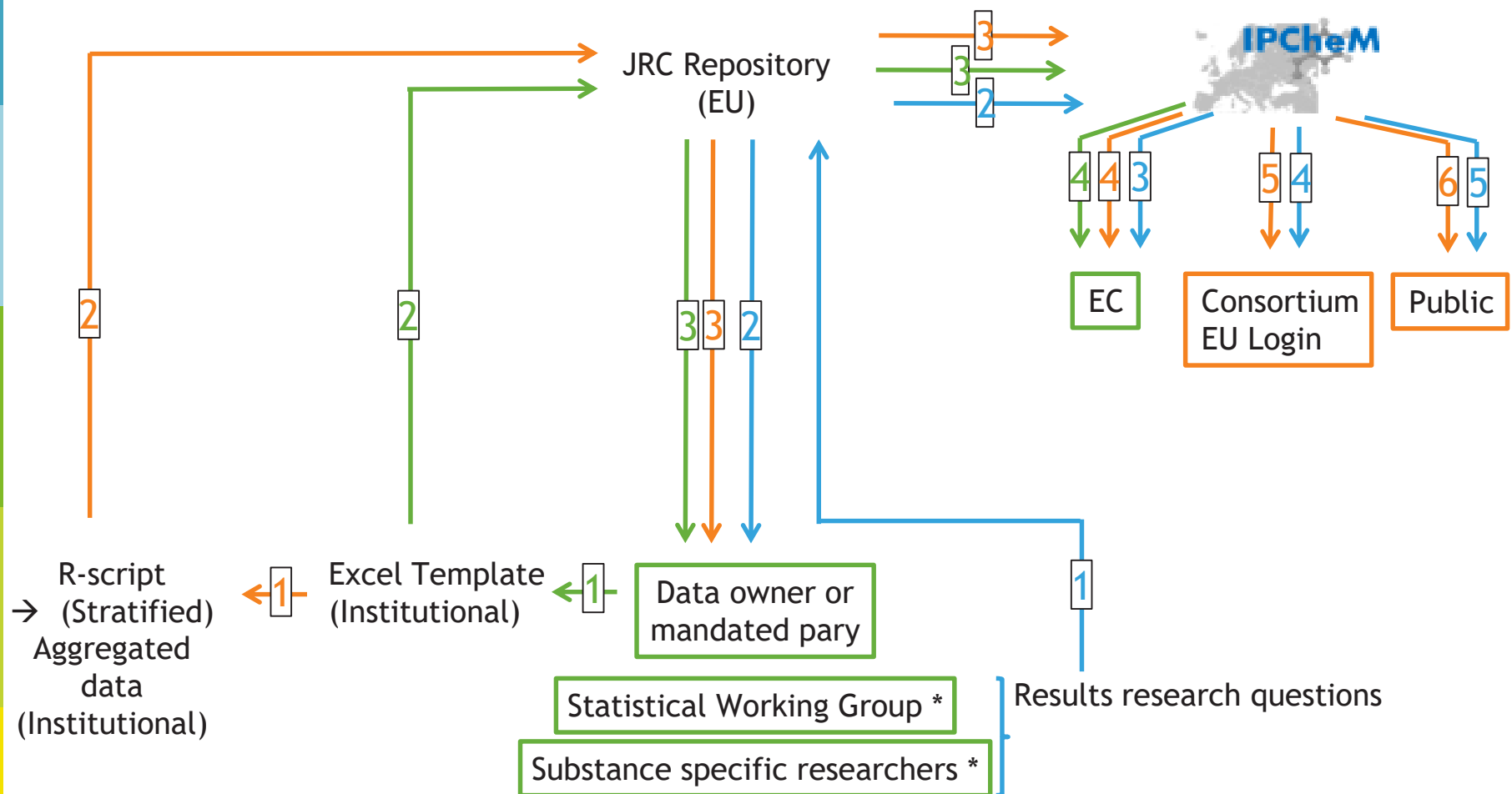
*→ This is a necessity to meet the objectives of HBM4EU!*

*Prior to generation of the data, the Data Owner/Data Provider shall confirm ethico-legal compliance of the study in which new data are generated; and fill out and sign the data transfer agreement.*

*→ Data generated with HBM4EU co-fund and accompanying variables are, by default, directly accessible for use within HBM4EU.*



# Data flow: New data generated with HBM4EU co-fund



Resolution of the data: Individual data

Aggregated data

Results research questions HBM4EU

\* Specified users may be granted/revoked specified access rights (e.g. to variables that are needed to answer research questions)

# Harmonisation of the data:

## Advantages

- Comparability of data
- Statistical analysis of pooled data collections
- Comparison between data collections
- Step forward in EU HBM (~Nhanes)
- ...

# Harmonisation of the (individual) data: HBM4EU codebook



HBM4EU  
codebook

Coding used to harmonize variables and data

# Harmonisation of the (individual) data: biomarkerlist



Biomarkerlist

In the biomarkerlist data providers have to indicate the metabolites they have measured and the matrix in which they were measured

→ Based on this information

→ Data management team will generate a  
→ Personalized data template

# Harmonisation of the (individual) data: Data template



HBM4EU data template

Data template



HBM4EU data template with example data

Data template  
example data

# Harmonisation of aggregated data: R-script

How will we obtain Harmonised aggregated data?

→ Data template for individual data = input file for an R-script to generate **harmonized aggregated data**

Output = **Descriptive statistics including:**

- ✓ N of subjects
- ✓ Arithmetic mean + SD
- ✓ Geometric mean + 95% CI
- ✓ percentiles with CI
- ✓ ...

→ Stratified for gender, age, BMI, smoking, level of education,...?

Open for input!

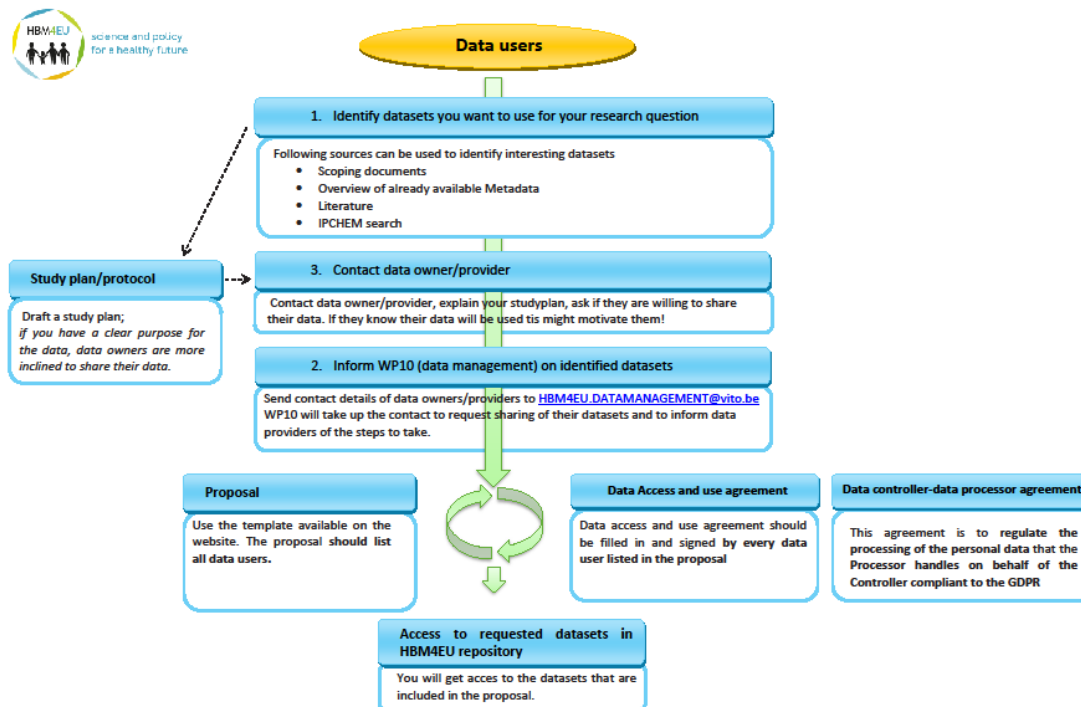
-> Tailored to the needs of the WP's within HBM4EU

# Advantageous of harmonized aggregated data

- We know how the data were processed!
- One way to deal with values  $< \text{LOD}$  or  $\text{LOQ}$
- Same descriptive statistics are calculated
- Comparability
- ...

➔ This will be the same for all datasets

# Requesting access to the data for data users





# Requesting access to the data for data users

1. Write a proposal

2. Sign data access and use agreement

3. Data controller –data processor agreement

# Requesting access to the data for data users

## 1. Proposal template



template

statistical protocol

- ✓ Define research Question/Purpose
- ✓ Studies/datasets that are needed
- ✓ Variables that are needed
- ✓ Records that are needed (e.g. sampling period)
- ✓ proposal for co-authorship

# Requesting access to the data for data users

- 2. Data access and use agreement



Data access and  
use agreement

✓ To be signed by all data users

**Signing this form the data user agrees to not abide these rules, accepts the terms and conditions of HBM4EU repository, and accepts the HBM4EU Data Policy.**

**Signing this form, the HBM4EU coordinator declares that the data will be made accessible from the HBM4EU repository to the data user for use for the purpose(s) outlined in the proposal.**

# Requesting access to the data for data users

## 3. Data controller – Data processor agreement

This document is in revision

The intention of this processor agreement is to regulate the processing of the personal data that the Processor handles on behalf of the Controller compliant to the GDPR

It will contain:

### 1. Subject matter of processor agreement

- ✓ categories of personal data considered
- ✓ purpose of the data processing
- ✓ Processing activities

->Only the personal data mentioned in the appendix of the agreement may be processed by data processor and for the specified purpose only!

# Requesting access to the data for data users

## 3. Data controller – Data processor agreement

### 2. Duration of the processing

The agreements applies as long as the data processor is processing data on behalf of the data controller in the agreed framework

### 3. Applicable legislation

Both Parties explicitly commit to comply with the requirements of the GDPR.

### 4. Controllers instructions

The Processor processes the personal data only on the documented instructions of the Controller set out in **Appendix 2** and shall not further process the personal data subject to this processor agreement in a manner which is incompatible with these instructions

# Requesting access to the data for data users

## 3. Data controller – Data processor agreement

5. Use of a sub-processor needs to be authorized by data controller

Processor shall not engage another sub-processor without specific authorization of the data controller

6. Confidentiality

The Processor commits himself to handle the personal data and its processing with utter confidentiality

7. Security measures

The Processor shall implement appropriate technical and organizational measures in such a manner that processing will meet the requirements of the GDPR and ensure the protection of the rights of the data subject.

# Requesting access to the data for data users

## 3. Data controller – Data processor agreement

### 8. Rights of the data subjects

Data processor has to assist the data controller in safeguarding data subjects rights

### 9. Assistance to data controller

In case of a data breach the data processor will notify the data controller

### 10. Transfer to third parties

The transfer to third parties, in any manner possible is prohibited, unless it's legally required or in case the Processor has obtained the explicit consent by the Controller to do so

# Requesting access to the data for data users

## 3. Data controller – Data processor agreement

### 11. Audit by controller

The Controller is entitled to evaluate the compliance with this processor agreement. He has the right to conduct an audit at any time on the location where the processing activities take place.

### 12. Liability

**The Controller** is responsible for obtaining the consent of the data subject for processing his or her personal data according to the provisions of the GDPR.

### 13. Intellectual property rights

All Intellectual Property Rights as regards to the personal data and as regards to the databases which contain these personal data are reserved to the Controller, unless otherwise contractually agreed upon between the Parties.



# Requesting access to the data for data users

## 3. Data controller – Data processor agreement

### 14. Termination of the contract

In the event of breach of this processor agreement or the GDPR, the Controller can instruct the Processor to stop further processing of the personal data with immediate effect.

This processor agreement can be terminated by both parties with a mutual period of notice of [fill in period].

# Thank you for your attention

If you have any questions about data management



Don't hesitate to contact the HBM4EU data management team!

[HBM4EU.datamanagement@vito.be](mailto:HBM4EU.datamanagement@vito.be)

All documents can be found on the website:

<https://www.hbm4eu.eu/data-management/>



VITO HBM4EU data management team:

[Sylvie.Remy@vito.be](mailto:Sylvie.Remy@vito.be)

[Eva.Govarts@vito.be](mailto:Eva.Govarts@vito.be)

[Liese.gilles@vito.be](mailto:Liese.gilles@vito.be)

[Greet.Schoeters@vito.be](mailto:Greet.Schoeters@vito.be)

[HBM4EU.datamanagement@vito.be](mailto:HBM4EU.datamanagement@vito.be)

[HBM4EU@vito.be](mailto:HBM4EU@vito.be)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733032.

21/06/2018